

Análisis Inteligente de Datos: Segundo Parcial

Claudio Sebastián Castillo

09 de mayo de 2022

ANOVA

Datos

Observaciones por grupo:

Se cumplen los supuestos para su implementación?

Anova

fit del modelo

coeficientes

p-value

F-value

Plot ANOVA

Conclusión

Testear homoscedasticidad

Test de Bartlett

sensibilidad al supuesto de normalidad

Testear normalidad

Testear normalidad analizando residuos

Anova y después: post-hoc

Tukey's Honest Significant Differences (HSD)

Cuando ANOVA no funciona: test de Kruskal-Wallis

ANOVA__multivariante

Analisis Discriminante Lineal (LDA)

Datos

Explorando discriminación por pares de variable

Homogeneidad de la Varianza: Histograma VariablesGrupo

Contraste de Normalidad Univariante Shapiro-Wilk

Contraste de Normalidad MultiVariante

Outliers

Test de Royston

Test de Henze-Zirkler

Contraste de Matriz de Covarianza

Estimación de parámetros de la función de densidad ($\hat{u}(X), E$) y cálculo de la función discriminante según aproximación de Fisher via `lda()`

Evaluación del error: Accuracy Table

Visualización de las clasificaciones

Analisis Discriminante Cuadrático (QDA) > falta de homocedasticidad/outliers LDA

Explorando discriminación por pares de variable

Contraste de Normalidad Univariante Shapiro-Wilk

Contraste de Normalidad MultiVariante

Outliers

Test de Royston

Test de Henze-Zirkler

Contraste de Matriz de Covarianza

Parámetros de la función de densidad función discriminante según aproximación de Fisher via `qda()`

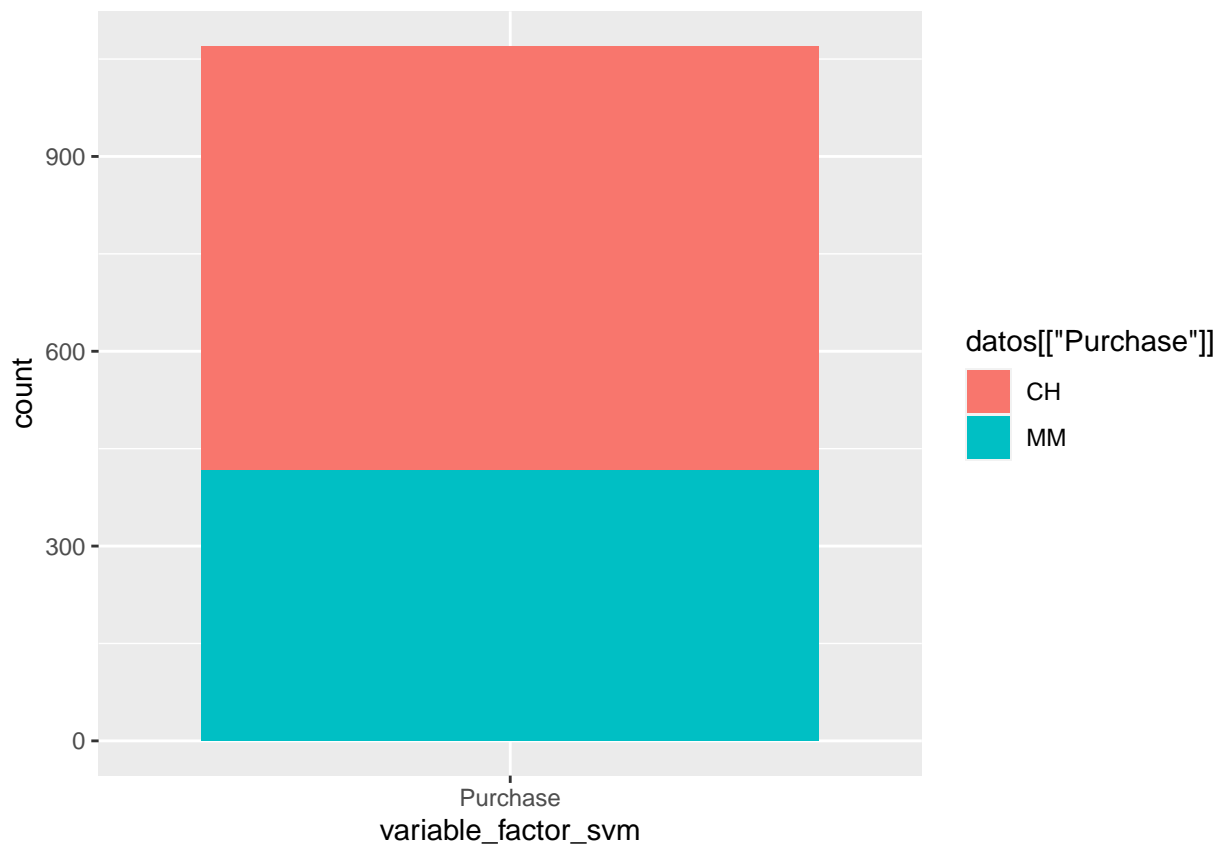
Evaluación del error: Accuracy Table

Visualización de las clasificaciones

Analisis Discriminante Cuadrático Robusto (RQDA) > falta normalidad

```
## $ Purchase      : Factor w/ 2 levels "CH","MM": 1 1 1 2 1 1 1 1 1 1 1 ...
## $ WeekofPurchase: num  237 239 245 227 228 230 232 234 235 238 ...
## $ StoreID       : num  1 1 1 1 7 7 7 7 7 7 ...
## $ PriceCH       : num  1.75 1.75 1.86 1.69 1.69 1.69 1.69 1.75 1.75 1.75 ...
## $ PriceMM       : num  1.99 1.99 2.09 1.69 1.69 1.99 1.99 1.99 1.99 1.99 ...
## $ DiscCH        : num  0 0 0.17 0 0 0 0 0 0 0 ...
## $ DiscMM        : num  0 0.3 0 0 0 0 0.4 0.4 0.4 0.4 ...
## $ SpecialCH     : num  0 0 0 0 0 0 1 1 0 0 ...
## $ SpecialMM     : num  0 1 0 0 0 1 1 0 0 0 ...
## $ LoyalCH       : num  0.5 0.6 0.68 0.4 0.957 ...
## $ SalePriceMM   : num  1.99 1.69 2.09 1.69 1.69 1.99 1.59 1.59 1.59 1.59 ...
## $ SalePriceCH   : num  1.75 1.75 1.69 1.69 1.69 1.69 1.69 1.75 1.75 1.75 ...
## $ PriceDiff     : num  0.24 -0.06 0.4 0 0 0.3 -0.1 -0.16 -0.16 -0.16 ...
## $ Store7        : num  1 1 1 1 2 2 2 2 2 2 ...
## $ PctDiscMM     : num  0 0.151 0 0 0 ...
## $ PctDiscCH     : num  0 0 0.0914 0 0 ...
## $ ListPriceDiff : num  0.24 0.24 0.23 0 0 0.3 0.3 0.24 0.24 0.24 ...
## $ STORE         : num  1 1 1 1 0 0 0 0 0 0 ...
```

Grfico datos



```
## [1] 857 18
```

```
## [1] 213 18
```

Busqueda de mejor hiperparametro C (coste)

```
##
## Parameter tuning of 'svm':
##
## - sampling method: 10-fold cross validation
##
## - best parameters:
##   cost
##     15
##
## - best performance: 0.1656772
##
## - Detailed performance results:
##   cost      error dispersion
## 1  0.001 0.2928454 0.04807128
## 2  0.010 0.1703010 0.04760141
## 3  0.100 0.1679754 0.04037149
## 4  1.000 0.1703010 0.04099747
## 5  5.000 0.1703010 0.03755497
## 6 10.000 0.1668263 0.03560699
## 7 15.000 0.1656772 0.03338668
## 8 20.000 0.1668399 0.03479758
```

Mejor modelo según hiperparametro

```
##
## Call:
## best.tune(method = svm, train.x = temp, train.y = datos_train[[{
##   {
##     variable_factor_svm
##   }
## ]]], ranges = list(cost = c(0.001, 0.01, 0.1, 1, 5, 10, 15, 20)),
##   kernel = "linear", scale = TRUE)
##
##
## Parameters:
##   SVM-Type:  C-classification
##   SVM-Kernel: linear
##     cost:  15
##
## Number of Support Vectors:  345
##
## ( 173 172 )
##
##
## Number of Classes:  2
##
## Levels:
##   CH MM
## [1]  1 15 24 28 32 45
```

Predicciones del Modelo

```
##          real
## prediccion CH  MM
##          CH 108 17
##          MM  22 66
## [1] "Observaciones de test mal clasificadas: 18.31 %"
```